

2016-09-28

Why Are There Developmental Stages in Language Learning? A Developmental Robotics Model of Language Development

Morse, AF

<http://hdl.handle.net/10026.1/8538>

10.1111/cogs.12390

Cognitive Science

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Why are there Developmental Stages in Language Learning? : A Developmental Robotics Model of Language Development

Anthony F. Morse, Angelo Cangelosi

University of Plymouth

Abstract

Most theories of learning would predict a gradual acquisition and refinement of skills as learning progresses, and while some highlight exponential growth this fails to explain why natural cognitive development typically progresses in stages. Models that do span multiple developmental stages typically have parameters to 'switch' between stages. We argue that by taking an embodied view, the interaction between learning mechanisms, the resulting behavior of the agent, and the opportunities for learning that the environment provides, can account for the stage-wise development of cognitive abilities. We summarize work relevant to this hypothesis and suggest two simple mechanisms that account for some developmental transitions; *neural readiness* focuses on changes in the neural substrate resulting from ongoing learning, and *perceptual readiness* focuses on the perceptual requirements for learning new tasks. Previous work has demonstrated these mechanisms in replications of a wide variety of infant language experiments, spanning multiple developmental stages. Here we piece this work together as a single model of ongoing learning with no parameter changes at all. The model, an instance of the Epigenetic Robotics Architecture (Morse 2010) embodied on the iCub humanoid robot, exhibits ongoing multi-stage development while learning pre-linguistic and then basic language skills.

Introduction

In normal development we progress from newborn infants to fully functioning adults passing through a number of ordered and protracted developmental stages, each revealing different cognitive capabilities

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

(what we can do, what we can not do, and the kinds of mistakes we are likely to make). Perhaps the most well known example of such stage-wise development comes from Piaget et al (Piaget, Cook, & Norton, 1952) who described four major stages of cognitive development: Sensorimotor (from birth to 18-24months) in which infants are only aware of what is immediately present; Preoperational (18-24months to age 7) in which children are able to think symbolically but their thinking remains egocentric; Concrete Operational (age 7 to 11) in which children recognize that their perspective and knowledge differs from others; and Formal Operational (age 11 +) in which reasoning gains abstract and hypothetical abilities. While the ages of transition between stages varies with each individual, the order of these stages remains rigidly fixed.

While Piaget focused on major stages, the acquisition of almost every cognitive skill also goes through protracted stages of development with widespread disruption (U-shaped performance) during the relatively short transitions between stages. The cause of staged development is most commonly attributed to maturational factors. Sometimes this is quite rational (for example; you can't learn to walk until your physical body has matured to the point of having enough strength to carry you) however in the case of intellectual abilities this form of explanation is assumed to appeal to changes in the neurological and hormonal machinery underpinning our cognitive abilities. This is normally assumed to be the result of major changes in the parameters or operation of that machinery (hence models that span developmental stages typically do so by varying the parameters of the underlying system to transition between stages), which in turn is also typically assumed to be genetically pre-specified. While we do not dispute that genetically specified operational changes do occur (for example; the massive reduction in neural interconnectivity in the early years of life (Huttenlocher, 1994), hormone changes during puberty (Sisk & Zehr, 2005), the ordering of myelination of cortical connections (Sowell, Peterson, Thompson, Welcome, Henkenius et al., 2003), etc.), we question how useful and how predictive these forms of explanations are beyond explaining a few specific well timed transitions.

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Taken out of context the staged development of most cognitive capabilities would seem in opposition to the general consensus that, mechanistically, learning is the gradual refinement (by whatever method) of statistical relations driven by experience. Indeed this depiction of learning would suggest that development should also be a gradual and incremental process, not one depicted by prolonged stages with relatively quick transitions (to clarify we are not saying that learning does not occur within each stage, rather that fundamental cognitive operations remain consistent within a stage and change between them). However, viewed in the context of an embodied agent, we suggest that the dynamic interaction between the environment, the agent, and its learning processes, provides a more useful and predictive account of some developmental stages. We also highlight that if stage transitions were solely the result of genetic pre-programming then there would not necessarily be cognitive indicators of how quickly or how slowly an individual would progress through the stages. While there are in fact many such examples, some of which we will discuss in what follows. In the remainder of this paper we will explore development specifically in the domain of language, however we suggest that the core ideas and developmental processes discussed are widely applicable to much of cognitive development.

Can Language be Learned?

Language learning is a particularly hard problem, so much so that many consider it an innate ability whereby a 'language of thought' is pre-specified by evolution, and only the translation to and from it remains to be learned (c.f. (Chomsky, 1995; Fodor, 1975, 2010; Pinker, 1994)). Clearly as the only animal known to naturalize language there must be some difference within us giving rise to this ability, but that thing need not be an innate language system (for example Pulvermüller (Pulvermüller, 2002, 2005) highlights additional pathways between the language centers and verbal motor areas present in humans which are absent from other primates, and Arbib (Arbib, 2005) discusses language readiness). The difficulty of the problem is clear if we take the perspective of the child as a passive observer viewing a cluttered scene while hearing a spoken word; to which feature or collection of features (visual or

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

otherwise), if any, should this word be associated? By contrast the dynamical systems perspective (Thelen & Smith, 1998) views the child as anything but passive; their attention is clearly focused and they are 'doing' (reaching, holding, banging, manipulating...) sometimes being physically lead by the caregiver (Zukow-Goldring & Arbib, 2007) and child directed speech is not simply directed at the child but also directed by the child's attention. Smith et al (Smith, Yu, & Pereira, 2011) go further highlighting just how dominant a held object is in the infants' field of view. From this perspective, in the sensorimotor stage the child is not really aware of all the perceptual clutter (the held object is simply occluding most of it), and spoken words often relate to what the child is currently doing / holding / attending, and so at least the learning of concrete item-based word-object and word-action mappings seems possible. That is to say, rather than passively viewing a scene cluttered with many objects, if the child's attention is specifically focused and much of the clutter is occluded, and the child directed language is similarly focused, then simple statistical learning should be enough to grasp many simple word-to-feature mappings.

Moving beyond simple word-object mappings Tomasello (Tomasello, 2000) further highlights that from a concrete item-based vocabulary children gradually (over many years) develop the ability to construct more abstract and adult-like linguistic constructions. This gradually increasing complexity of language presents a significant challenge to the hypothesis that language is innate. Herein we describe an approach to model and explain this learning process, and attempt to understand its staged developmental time-course. Rather than present new experiments here, this paper sets out to piece together existing work, with reference, highlighting how a fairly simple modeling approach can account for a wide variety of experimental data in a single model displaying developmental stages without any system or parameter changes.

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

The Developmental Time-Course of Language Learning

From birth, and perhaps before, infants are sensitive to linguistic utterances, displaying sensitivity to changes in single phonemes even when those phonemes are in a non-native language. For example Eimas et al (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) played speech sounds to 1 month old and 4 month old infants, habituating them to one sound and measuring discrimination by an increased conditioned response rate to a second sound. Stager et al (Stager & Werker, 1997) similarly tested 8 month old infants in the 'switch task' in which they are first habituated to one image paired with a multi-phoneme construction (e.g. 'lif'), and then to a second image paired with a second phoneme construction. Finally the infant is presented with the original image paired with either the original phoneme construction or a new one varied by one phoneme (e.g. changing 'lif' to 'rif') or by multiple phonemes (e.g. changing 'lif' to 'neem') and their looking time measured. For 8 months old infants, looking time is significantly greater for both forms of changed phoneme constructions but not for the original pairing, suggesting that they are sensitive to the phoneme manipulations in both native and non-native language phonetics. Herein we refer to this as the *universal phonetic discrimination stage* (0-8 months).

By 12 months old, however, infants are significantly less sensitive to the single phoneme changes when paired with images as in the switch task, and it is not until 14-17 months that this sensitivity is regained, which, crucially is only regained for native language phonetics. Maye et al (Maye, Werker, & Gerken, 2002) highlights that the understood vocabulary size is an indicator of when this transition will occur, a result that also challenges the hypothesis that this transition results from genetically specified maturation, suggesting rather that the infants' own cognitive abilities are somehow driving this transition. This marks a new developmental stage in which the child not only grasps the target of utterances but also begins to produce them (first words typically occur around 10 months of age). Herein we will refer to this second developmental stage as *perceptual reorganization* (~14 – ~20 months).

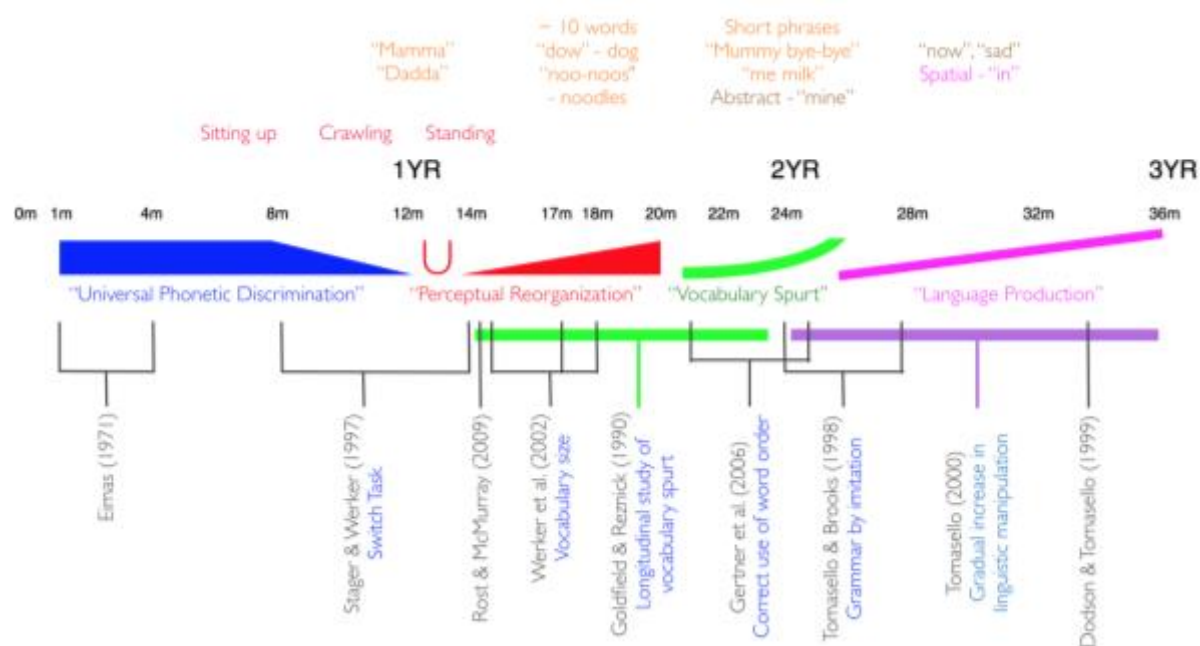
This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Once in this stage the infants' vocabulary size grows as new words are learned slowly and painstakingly, requiring many exposures to each new word before it is reliably grasped. By around 18 months the average child can produce about 10 concrete item-specific words but can understand many more. From around 20 months of age however, the size of their production vocabulary begins to grow at an exponential rate (Goldfield & Reznick, 1990). While the existence of a vocabulary spurt is contested (Ganger & Brent, 2004) debate is focused on whether word acquisition rates increase and then level off or whether they continue increasing. Given that there is some delay before spatial, emotional, or temporal concepts are used, one currently untested possibility is that several 'spurts' occur in different concept areas overlapping to varying degrees, either way what is important to this review is that the vocabulary acquisition rates clearly undergo a significant change around this age. McMurray (McMurray, 2007) demonstrates that this vocabulary explosion can be explained by parallel learning and the statistical distribution of word difficulty across the language; if we assume that it takes a certain number of exposures to a word before it is grasped, and that different words occur with different frequency then we should expect an exponential growth pattern. While this clearly is a major factor in the 'vocabulary spurt', Goldfield and Reznick's longitudinal study of word learning (Goldfield & Reznick, 1990) and various fast mapping experiments (e.g. (Horst & Samuelson, 2008)) highlight that from the onset of the vocabulary spurt new words can be learned from single exposures in ambiguous conditions (this is unexplained by McMurray). Gertner et al (Gertner, Fisher, & Eisengart, 2006) also demonstrate that, during the vocabulary spurt, as new words are learned they are understood and produced in the correct order (e.g. 'the bunny is gorping the duck' Vs 'the duck is gorping the bunny' (here 'gorping' is a novel verb introduced in the experiment)), showing that word order cues are being learned at the same time as the words themselves. Due to these new abilities (learning words from single exposures in ambiguous context, and learning word order cues) we will refer to this as the *vocabulary spurt stage* (20 – 24months), though it may equally be seen as a transition to the next stage. Children in the vocabulary

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

spurt stage typically produce short phrases (e.g. 'mummy bye bye', 'me milk') and start to use more abstract / less concrete concepts such as 'mine', but are not yet producing utterances referring to spatial, emotional, or temporal concepts.

Following the vocabulary spurt the child's language gradually becomes more abstract, with spatial, emotional, and temporal utterances typically first occurring around 30 months. This slow progression toward longer constructions and more adult-like language seems to progress almost linearly, for example; Tomasello (Tomasello, 2000) reviews a number of experiments in which children from 24 months to 8 years demonstrate a linearly increasing tendency (with age) to use a novel verb in a new transitive utterance when that novel verb has only be heard in an intransitive utterance. Such manipulation of linguistic structure is the hallmark of language and separates it from the mere mapping of words to concepts, and thus this is the final stage of language development we will consider for now. Herein we refer to this as the *language production stage* (24 months +). These four developmental stages are depicted on a timeline in Figure 1 below.



This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Figure 1 Showing the early developmental stages of language production from birth (left) to 3 years of age (right), each stage is depicted by a different color/shape across the middle of the figure. At the top we depict onset stages of language production, and mark major motor developmental milestones. Across the bottom we mark some of the references discussed in the text showing the ages at which their results were obtained.

Modeling The Developmental Time Course of Language Learning

How Can Developmental Robotics Inform Cognitive Science

Before psychology and the cognitive sciences can mature, the accumulation of scientific knowledge must first fit, support, form, and revise wider theoretical perspectives and paradigms. To this end hugely important experiments are conducted, and theoretical positions are continually revised and developed and yet there often remains a gap between the two. Theories, while informed by experimental data and discoveries, often lack the detail to make specific claims and predictions. Modeling has the potential to provide a bridge between theory and experimental data in that the modeler is forced to be explicit about the specific details required to make the theory work. An implemented simulation is therefore not only demonstrably able to replicate the data being accounted for, but is significantly more theoretically tight than the abstract flow charts or boxologies we commonly see. While modeling has been hugely influential in cognitive science, all too often the various models focus on specific tasks and though supposedly implementations of the same theoretical position (even when using the same methodology, such as a production system) they do not themselves appear compatible. It is not then that we just need modeling to provide greater detail in proposed theories but we also need to demonstrate integration between or across phenomena. We need the same model (not just the same methodology or approach) to account for a variety of phenomena and thus to link them to the same underlying mechanisms and dynamics (Morse, DeGreeff, Belpeame, & Cangelosi, 2010; Morse, Herrera, Clowes, Montebelli, & Ziemke, 2010; Newell, 1990). Developmental Robotics takes this a step further, forcing not only the integration of

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

cognitive faculties but perceptual and behavioral abilities too, as an autonomous robot that cannot integrate all the way from sensory stimuli to behavior output is simply insufficient. Using a humanoid robot like iCub means that we can replicate psychology experiments in unprecedented detail by interacting with iCub in exactly the same experimental setup as used with human participants, even using the same physical stimuli. Obviously there are important differences between iCub and human participants and we are the first to admit that iCub's sensorium and motor systems are not that of a child. However the demonstration of a single detailed mechanism in ongoing development, interacting with experimenters who replicate a variety of child experiments across multiple developmental stages is a big step in scaling up this approach, and one which can make testable predictions and provide detail to developing theories of cognition.

Previous work has demonstrated this architecture replicating a variety of psychology experiments, however the aim of this paper is to put these pieces together, herein we summarize that work with reference to the original publications so that the reader can find the details.

The Epigenetic Robotics Architecture (ERA)

Originally proposed in 2010, the Epigenetic Robotics Architecture (ERA) (Morse, DeGreeff, Belpeame, & Cangelosi, 2010) is a simple but scalable modeling approach for integrating phenomena in ongoing development. ERA retains transparency, in that its knowledge is organized in recognizable conceptual structures that emerge through ongoing interaction with its environment. While we present a specific implementation (details in Appendix), much of ERA is about the interactions between parts of the model and so long as the required functionality of the parts remains consistent then the details of this or that classification algorithm are not the focus of interest. To this end we use a fairly simple neural network classification method, self-organizing maps, to classify data coming from different modalities or different sources (object colour, object shape attributes, phonemes, words, body posture (joint angles), actions). We acknowledge alternatives here such as Dynamic Field Theory (Erlhagen & Schöner, 2002) or Deep

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Belief Networks (Hinton, Osindero, & Teh, 2006), and suggest their substitution would likely not affect the dynamics on which we base most of the explanatory aspects of the model, however the initial transition requires the modified SOM detailed in the Appendix). This provides a set of classifiable features from sensory and motor data, which we then associate via learning (normalized Hebbian learning) in a structured way to provide an associative memory not unlike the hand wired Interactive Activation and Competition (IAC) models (McClelland & Rumelhart, 1981) that have been used to explain various cognitive phenomena in early connectionism. Unlike IAC however our structure emerges through ongoing interaction and we can now use this to explore development and the effects of specific interactive experiences on the developing model. For a more detailed discussion of the model including equations and parameters used see Appendix and Morse and colleagues, 2010. (Morse, DeGreeff, Belpeame, & Cangelosi, 2010). The model schematic is shown in Figure 2 below. How the ERA model accounts for psychological data in replications of experiments across the developmental stages, and how it transitions between them is discussed in more detail in the remainder of this paper.

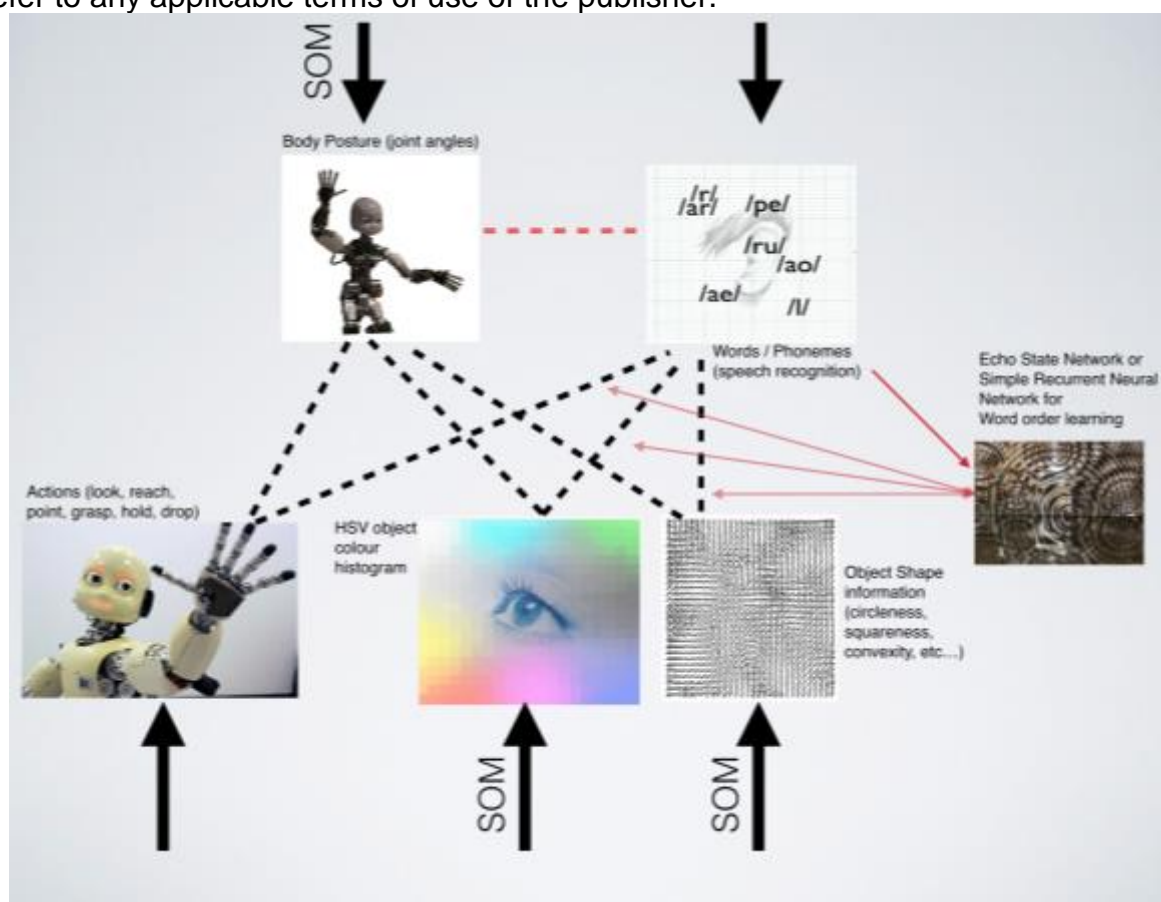


Figure 2 Schematic representation of the ERA architecture. Each modality is driven by sensory input or motor activity from the iCub robot, classified by a modified Self-Organizing Map, and associated (black dashed lines) in a structured way to specific other SOM activities. In addition we show the word order learning mechanism, which dynamically adjusts the learning rates on connections between specific SOMs.

Modeling The Universal Phonetic Discrimination Stage

To begin modeling this time course we start at the Universal Phonetic Discrimination stage (0-8 months), where we need not only a method to discriminate previously experienced phonemes but one that can also discriminate novel phonemes in-order to replicate sensitivity to non-native language. Herein we will use Self-Organizing Maps (SOMs) to learn and discriminate between input patterns. While various studies show that sensitivity to phoneme changes alone remains high across this period, Stager & Werker (Stager & Werker, 1997) show that the pairing of an image with a phonetic utterance causes disruption from 8-months to 14-months recovering only for native language utterances. To this end we need a simple multi-modal model and we can achieve this by providing 2 SOM's, one driven by the phonetic signal and the other by a visual signal (in the simplest example we use color). Next we use a simple

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Hebbian (Hebb, 1952) learning rule to form associations between units in these 2 SOMs, however we further adapt the SOM learning algorithm so that this multi-modal priming (spread of activation via these associations) is influential on the SOM development.

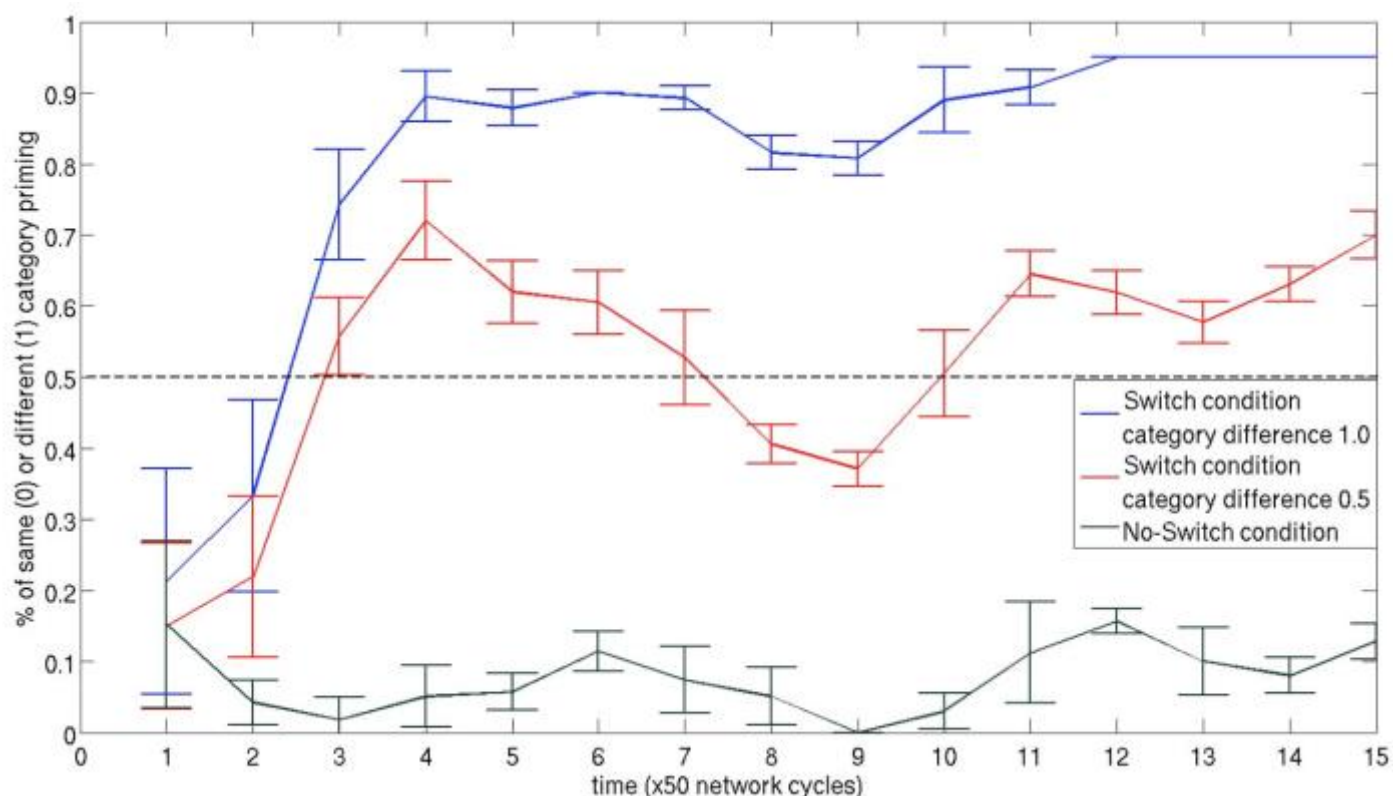


Figure 3 Replicating Stager & Werker's switch task experiment, this figure shows the mean category judgment for each 50 consecutive trials across 20 individuals for 3 different conditions; Switch trial with a large difference (1), Switch trial with a small difference (0.5), and No-Switch trial, corresponding to a large phonetic change (e.g. 'lif' to 'neem'), a single phoneme change (e.g. 'lif' to 'rif'), and no change ('lif' remains as 'lif'). Large changes are consistently recognized (top of graph), no-change is consistently recognized as familiar (bottom of graph), but sensitivity to single phoneme changes (middle of graph) undergoes a U-shaped disruption between 7 (350 network cycles) and 10 (500 network cycles).

From Figure 3 above we can see that presentation of an image with its associated phonetic utterance (the no-switch condition) is consistently recognized as familiar (bottom line of the graph) (i.e. not new / novel / different), and that following an early period of organization (while the SOM learns), an image paired with a very novel phonetic utterance remains recognized as unfamiliar (top line of the graph) (i.e. new / novel / different). What is interesting about this experiment is that an image paired with a slightly changed phonetic utterance (the mid line in Figure 3), following initial organization of the SOM, is recognized for a while as novel (indicating sensitivity to the phonetic change) but then goes into decline

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

(indicating a lack of sensitivity to the change) before finally recovering and being recognized as novel

again. Crucially this recovery is only for experienced phonetics while sensitivity to novel phonetic –

image combinations are never regained.

In the initial phase of this experiment, both SOMs (initially randomly configured) go through a significant re-organization as learning progresses. This has consequences for the learning of associations between the SOMs, as despite the associative learning being ongoing throughout, the fact that the same stimulus will elicit a different response from units over time means that there is no consistency of image and speech responses to be learned. However once the SOMs have achieved a somewhat stable organization and responses are more consistent, only then will the associative links have the chance to develop more strongly. To be absolutely clear both learning algorithms (the SOM and the Hebbian learning) are active from the outset, it is not that we need to detect that the SOM has stabilized before switching on the hebbian learning, rather that both are active throughout and the stabilizing of the SOM brings about the conditions necessary for the hebbian learning to gain purchase on the relationships between SOM calssifications. As these learned relationships gain in strength and influence they impact on the developing SOM, initially disrupting it but also biasing its ongoing learning toward categorical responses with predictive utility, i.e. biasing categorization in favor of useful classifications to the detriment of non-useful ones (i.e. native language over non-native language). For full details of this experiment see Morse et al (Morse, Belpaeme, Cangelosi, & Floccia, 2010)

This is a prime example of neural readiness, as the two learning processes are interacting. In this example it is not until one learning process (here the SOM learning algorithm) has produced strong and somewhat stable connections that the second learning process (here the associative learning algorithm) can gain purchase and have any significant effect. However once both become influential they initially

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

disrupt each other until some equilibrium is reached. This is the transition from the universal phonetic recognition stage into the perceptual reorganization stage.

Modeling The Perceptual Reorganization Stage

From this point on the model is capable of cross-situational learning (Belpaeme & Morse, 2012), in that exposure to multiple visual stimuli paired with an utterance strengthens the association between them such that multiple exposures to varying sets of features paired with the same utterance will gradually refine the associations until the actual target feature of the utterance is strongly associated to it.

To illustrate the first developmental stage we used a very simple model of only 2 SOMs associated to each other, however to demonstrate further abilities we need to scale up the model, retaining the same structure but simply adding more SOMs driven by shape information, body posture information, and complete words. Rather than associating every SOM to every other SOM we simply associate every SOM to the word and body posture SOMs allowing indirect spreading of activity between unassociated SOMs via these 'hubs'. Again full details and a rationale for this architecture can be found in Morse et al (Morse, DeGreeff, Belpaeme, & Cangelosi, 2010). While it may appear that we are making significant changes to the model part way through its development, the new more complex model can be started from a random starting point (i.e. SOM weights randomized and associations re-set to zero), and will display the same developmental behavior previously discussed; starting in the universal phoneme recognition stage, and following some period of learning transition into the perceptual reorganization stage. The additional complexity simply allows us to demonstrate more complex interactive behavior using the iCub humanoid robot.

As word-object mapping is learned, slowly and cross-situationally, in the perceptual reorganization stage, several new infant abilities can be observed. The first is fast-mapping, where novel words are mapped to novel items in the presence of multiple known items (sometimes referred to as mutual

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

exclusivity), and the second is the simultaneous learning of word order with words. We will start with fast-mapping.

In the presence of multiple objects, multiple SOM features are simultaneously active in each SOM, and as some learning has already happened by this point those features will be associated with various words, and those words, which are significantly primed, will suppress other words that are not significantly primed. This is in the nature of a properly implemented neural SOM, however many implementations simply ignore inhibitory links between units in a SOM replacing this step with a computational 'winner takes all' algorithm, herein we require the inhibitory connections. Given this property, the strong external priming of a heard word not only suppresses (via within SOM inhibitory connections) other words (that may be primed by association to visual features), but in effect slightly suppresses their associated features as well. To be clear, because the activation of other, non-heard, words falls below zero, they exert an inhibitory influence on anything positively associated with them.

The effect is that a novel word slightly suppresses features previously associated with other known words and so subsequent association between the new word and the full set of features present is stronger for the unknown features and so the model is biased toward linking the new word to the novel object. While not completely intuitive this is simply a description of the natural dynamics of a whole class of multi-layered bi-directional associative networks with inhibitory pools (in this case the SOMs) and excitatory connections between pools (for the simplest examples see (Burton, Young, Bruce, Johnston, & Ellis, 1991; McClelland & Rumelhart, 1981)).

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

To replicate specific fast-mapping experiments we first have to establish a set of known objects; this is achieved by repeatedly presenting a series of objects, over multiple interactions, to iCub and naming them each time. This is repeated until iCub is able to reliably name them all, learned by slow cross-situational learning. Once this set has been established we are able to begin the experiment. Replicating the fast-mapping experiments (Twomey, Morse, Cangelosi, & Horst, 2014) we present combinations of multiple objects (either all known, or including a single novel object) and ask iCub where one of them is (verbally asking by the name of the object) (see Figure 4 below).



Figure 4 iCub responding the question 'where is the dax' (dax is the novel word) by pointing at the novel object

Several variations of this basic scenario have been run to examine different behavior and compared with child results from multiple experiments. For example, performance at fast mapping declines rapidly when more distractors (known objects) are present when asked to find an X (where X is novel word). Figure 5 below highlights the close match between the robot and child data for this task (see Twomey et al (Twomey, Morse, Cangelosi, & Horst, 2014) for full details of this experiment).

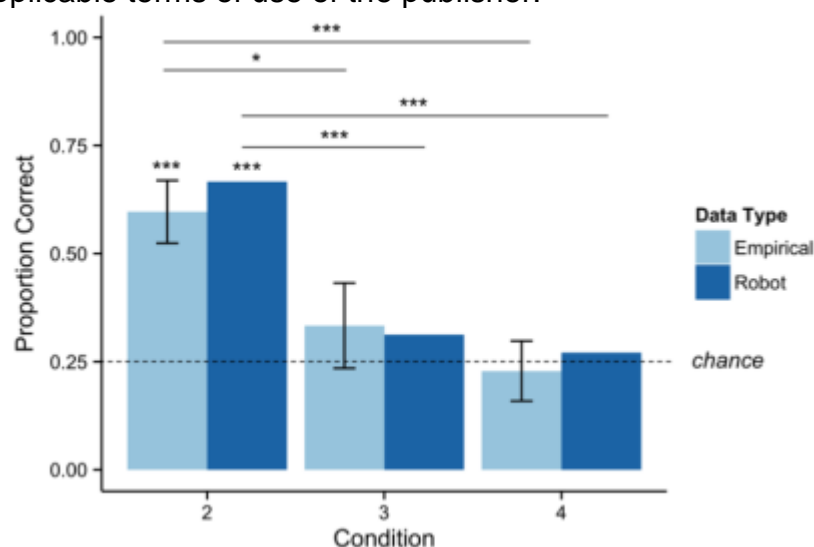


Figure 5 Proportion correct of children's (light blue bars) and the model's correct choices (dark blue bars) at test, for fast-mapping a novel object to a novel word in the presence of 2, 3, & 4 known distractors. *** $p < .001$, ** $p < .01$, * $p < .05$.

Similarly retention and extension have been explored following presentations of narrow or wide variance in the novel object on subsequent naming trails leading to differences in generalization and category formation (Twomey, Morse, Cangelosi, & Horst, 2014), and recent work (Horst & Morse not yet published, see Figure 6 for a summary of results) has explored what happens when features such as color are held constant across known and novel objects, in all cases demonstrating a very close fit between this model and the child data in a variety of fast-mapping experiments.

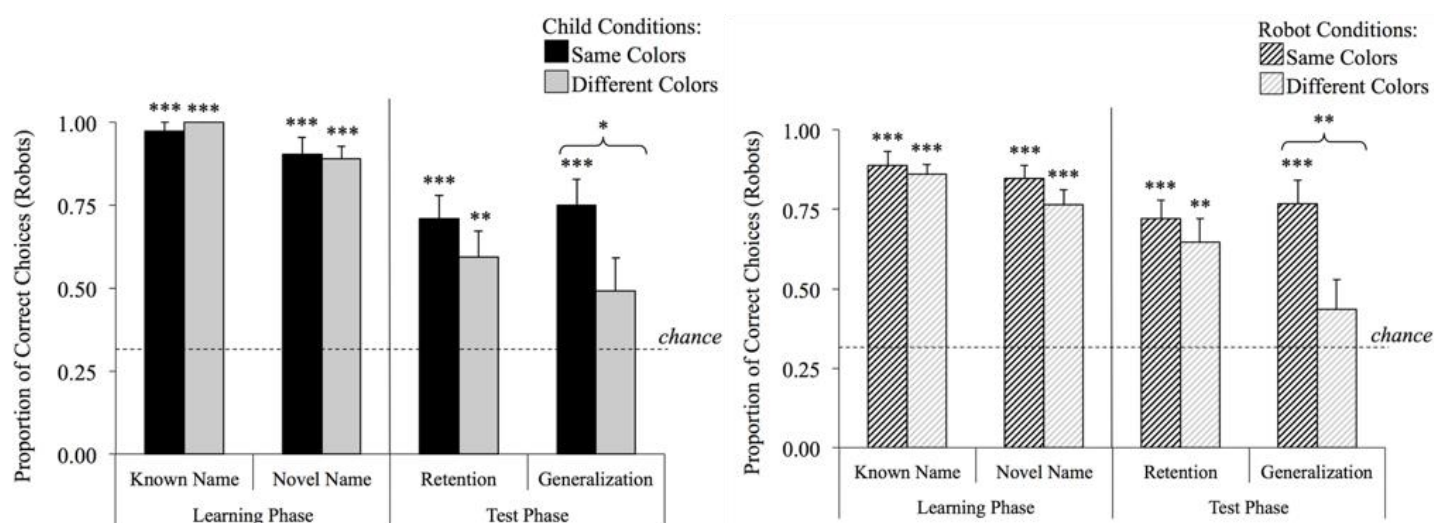


Figure 6 Comparison between child data (left) and robot data (right) in a fast-mapping experiment in which performance is measured when either known and novel objects vary in only shape, or vary in both shape and color. In the learning phase

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

performance is measured for selection of the correct object (known or unknown). In the test phase performance is measured for selection from combinations of novel objects that are either the same (retention) or similar (generalization) to the original novel objects used in the learning phase.

Modeling The Vocabulary Spurt

The next phenomenon apparent in this stage is the simultaneous learning of words with word order cues (Gertner, Fisher, & Eisengart, 2006). Given that we now have multiple SOM modalities which to some extent map onto word categories (color can be an adjective, action a verb, and shape somewhat more strongly related to the object / noun), we can learn the relation between the position of a word in an utterance and to which SOM that word is most strongly associated to. Again there are many ways this could be achieved, we have tested with offline training of a Simple Recurrent Neural Network (SRNN) (Elman, 1991), and ongoing work is testing this learning online using an Echo State Network (ESN) (Jaeger, 2001) with potential to learn many other 'grammatical' features of utterance organization (Frank, 2006; Tong, Bickett, Christiansen, & Cottrell, 2007). Regardless of the method employed the information about which SOM is most likely to be associated with each word position in a given utterance can be used to dynamically modify the associative learning rate for those connections, amplifying or subduing its effect on the connections to each of the SOMs.

In combination, fast-mapping and biasing learning with word positions, provide the model with the ability to learn novel word-object mapping in ambiguous situations, and from a single exposure (see Figure 7 for an example). We suggest that this ability is a key component of the vocabulary spurt and plan to further establish this in future work. With this in place, and having already interacted with iCub through the previous experiments, we can now demonstrate the learning of new word-object mappings in ambiguous context. We start by showing iCub a novel item (novel in both colour and shape) such as a leaf and saying 'this is a green leaf' (the words 'green' and 'leaf' are new to iCub). The model will strengthen connections between the SOM units classifying the words and the color green, and the shape of the leaf, BUT the word 'green' will be more strongly associated with the color and the word 'leaf' will

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

be more strongly associated with the shape, both due to previous grammar learning predicting their association to particular SOMs based on their position in the sentence. So far this is fairly straight forward however we next place two novel objects in iCub's field of view, a green carton and a purple carton, and ask where the purple carton is (again completely new words for iCub). Word order prediction suggests that the new word 'purple' is most likely a colour, and the colour of the green carton is already associated with the word 'green' and so through fast mapping the salience of the purple carton is higher than that of the green carton and so the purple carton is selected. See Figure 7 for a depiction of this experiment. This is new result presented here for the first time and will be the basis of extended experiments in the future designed to quantify and qualify this result formally.



Figure 7 Showing an example timeline of the model in the vocabulary spurt stage. (Left) iCub is first shown a novel item, the green leaf and told to 'look at the green leaf'. (Middle) iCub is then shown two novel items, the green and purple cartons, and asked 'where is the purple carton'. (Right) iCub then orients to and reaches for the purple carton. Prior to this iCub has never seen a green leaf or a carton or anything else that is purple, and has never heard the words 'carton', 'purple', 'green', or 'leaf'.

At the outset of this paper we suggested that two mechanisms, neural readiness and perceptual readiness, could be usefully employed to help explain and understand why development progresses in stages. The learning of word order provides a good example of perceptual readiness in action, in that before any word order effects can be learned, the model must first have grasped a small vocabulary of word - concrete-item mappings spanning the different modalities of information. It must have learned a combination of color words, shape words, and action words. Thus it must be able to perceive and prime or predict those words before it can learn their relative utterance positions, and then generalize these positions to new words. Again to emphasize the point, it is not that we wait for sufficient words to have

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

been bound to their visual targets before we switch on the learning, rather that the learning is always going on BUT the informational requirements for learning word position cues are not initially met and so the learning process makes no headway. Only once a set of word-target mappings have been acquired are the informational requirements for word order learning met, and generalized to novel words in the same utterance positions.

Modeling The Language Production Stage

Demonstrating the Language Production stage is ongoing work and replicating the specific findings of Tomasello (Tomasello, 2000) and others will require more than just word order cues, however, the model described can already use the learned word order cues in reverse. For example on seeing an object, features in multiple SOMs are activated (the objects' color profile and shape information), each of which may prime, by previously learned associations, different words such as 'red' and 'car'. The question is in what order should these words be spoken by iCub - 'red car' or 'car red'?

In this example we can assume that only 2 words are strongly primed (though the same method will also work for longer utterances) so we can provide the 2 primed words to the word order network and we find that the first word amplifies the color SOM which is where we find 'red', and the second word amplifies the shape SOM which contains 'car'. Thus the model will say 'red car', and will not say 'car red'.

In effect this is mining the model for information and does not naturally fall out of the inherent dynamics, however current work replacing the SRNN with an ESN is anticipated to change this. The ESN is a bi-directional network such that features can stimulate words and words can stimulate features. Combined with the abilities of ESNs to capture more complex grammar-like rules we hope to be replicating various psychology results from this fourth developmental stage in the near future.

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Final Summary

In each developmental stage we have focused on different parts of the model and built up the description but given this full version of the model we can now go back to the start, resetting associative connections and randomizing the SOM weights, so that we have an infant again. With the full model, and without any external parameter changing we can now interact with iCub, showing it various objects, talking to it, and inviting it to interact with those objects (picking them up, pointing at them, telling us what it thinks they are, showing us where the X is etc.). At first progress is very slow, iCub doesn't know much, though it reacts to phoneme changes (using experienced phonemes, and completely new ones). After a while iCub's sensitivity to small phoneme changes when combined with visual stimuli reduces briefly before eventually coming back only for experienced phonemes. Around this point iCub is able to point at the correct object if it has had multiple experiences of that object being named. This continues for a while and iCub is able to do fast-mapping experiments as its repertoire of known words grows slowly. iCub's speech is initially disordered (e.g. 'car pickup red') but soon becomes ordered, even for newly acquired words. Once this has happened iCub starts to learn new words relatively quickly, sometimes from single exposures when multiple possible target mappings are present. We believe this demonstrates the same progression through developmental stages that children go through, and have demonstrated the close fit to child data in multiple experiments across these stages.

Just as infants progress through a series of clear developmental stages in early language development, the model progresses through the same series of developmental stages as direct result of ongoing learning and experience during physical interaction and without parameter changes.

Transition between the stages is all about first meeting the requirements (informational or neural) for the next set of competences to emerge. If we consider the transition from Universal Phoneme Recognition to Perceptual Reorganization, this is about neural readiness. A set of weights or connections

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

(in the SOMs) must become stable before a response is consistent enough for multimodal associations to reliably gain strength and hence influence the further development of the SOM by priming. This principle of refining the set of detectable features to maximize their predictive utility is of importance to more general theories of cognitive function, and is consistent with increasing neuroscientific support of the idea that prediction may be a universal principle of brain operation (Bar, 2007, 2011). Theories that attempt to bridge the gap between basic prediction and cognition, such as sensorimotor perception (Gallese & Lakoff, 2005; Hesslow, 2002; Noe, 2004; O'Regan, 2011; O'Regan & Noe, 2001) have a tendency to focus on sensorimotor knowledge as learned associations between a fixed set of detectable features, however the approach herein additionally provides an example of how to refine the set of detectable features to maximize their predictive utility. We suggest that this may be an important feature typically overlooked in sensorimotor theories.

The next transition, to the Vocabulary Spurt, requires learning the word order cues, but this has an informational requirement that words and their target modalities must first be mapped (at least for a sub-set of the words) and this takes time with slow cross-situational learning. However once word order cues are learned, then combined with fast-mapping, we can use completely new words for example; 'look at the dax doff' in the presence of multiple objects, and using the word order ('dax' must be an attribute of the object 'doff'), and fast-mapping (that's a 'car', that other thing is 'green', so it must be this 3rd object), iCub relates 'dax' to the color profile of the 3rd object, and 'doff' to its shape.

At this point iCub's language skills remain impoverished by comparison to typical 2yr old children perhaps being better characterized as word-object mapping, as language implies a far greater richness of lexicon and word manipulation skills. However the sensory and motor abilities of iCub used in these experiments are also impoverished, by providing a richer set of skills and sensory sensitivities (perhaps through deep learning) we would anticipate richer language to result. So can language be learned? We

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

argue that in these various experiments we have shown that many of the features of early language learning, including the emergence of developmental stages, can be learned, and that we can further explain the necessity of ordered stages of development without recourse to genetic or physical maturational events.

In this paper, we have discussed specific models that replicate data from existing experiments. However, our account of developmental progression is more general than those studies. The explanation we propose is based on the dynamic neural and informational pre-requisites for the learning of behavioral capacities that define each stage. This is a stronger account than maturational as for example it explains why vocabulary size is a good predictor of when an individual child will transition to the second developmental stage, why word order recognition abilities predict the onset of a vocabulary spurt etc.

This paper has therefore provided a summary of a research project demonstrating how a single model has replicated a variety of psychological phenomena across developmental stages in language.

Acknowledgements

This work has been supported by the EU FP7 projects ITALK (no. 214668) and Poeticon++ (no. 288382). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

Appendix – The ERA model

The Epigenetic Robotics Architecture (ERA) is a hybrid neural network supported by various pre-processing stages. The camera feed from iCub is first processed to identify and mask objects based on saturation levels, each masked object is processed to provide a normalized vector histogram of the HSV color values of the pixels within the mask, this provides 32 input values to the color SOM. The same mask

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

is processed for shape information (using the same method as described here (Gonçalves, Abrantes, Saponaro, Jamone, & Bernardino, 2014) for convexity, circleness, squareness, elongation etc.) and this information provides input to the shape SOM. Joint angles from the robot are read directly and provide input to the Body Posture SOM. A set of actions are pre-defined to make iCub point, pick-up, and drop, objects when a location is specified. Dragon Dictate is used to provide free speech-to-text and then each new word spawns a new neuron in the word map, while previously encountered words stimulate their associated neuron. Mel-frequency coefficients drive the phoneme SOM (which is only able to reliably discriminate a few specific vowel sounds). The Body Posture SOM, and the word map, are then fully connected to each of the other SOMs with weights at zero under the influence of normalized hebbian learning. For each camera frame, each object is scanned in turn to provide sequential input to color and shape SOMs, and then the states driven by each object are superimposed (i.e. multiple 'winners' may be active), and interact (inhibiting each other but with fixed external input on). The object with features most strongly active at any point becomes the 'salient' object and iCub turns its head and eyes to look at that object, thus when speech primes an object, it becomes salient and iCub turns to look at it. Actions are hard-wired to specific words in the word map such that when heard they trigger the action, which will be directed at the object iCub is looking at.

The SRNN version of word order recognition is a standard SRNN with 10 binary inputs, and 3 outputs. For each sentence that is heard, a series of training inputs are generated as follows:

For each input pattern generated the length of the sentence is on, and each separate input pattern contains one word position on. The corresponding training target for each pattern is the relative priming activity levels of the three SOMs (action, color, shape) if that word was heard in isolation. Every 10 sentences, the SRNN is batch trained with standard backprop on the cumulated training data. For each word of any sentence, the output of the SRNN for that words position is used to dynamically modify the learning rate for all connections between that word and the various SOMs.

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Equation 1: Initial direct activation of SOM units

$$DirA_j = \sqrt{\sum_{i=0}^{i=n} (v_i - w_{ij})^2}$$

Where $DirA_j$ is the resulting activity of each node in the map following a forward pass of the SOM, v_i is an input, and w_{ij} is the weight between that input and the current node. The winning node is the node with the smallest value for $DirA_i$

Equation 2: Initial indirect activation of SOM units

$$IndA_j = \sum_{i=0}^{i=n} x_i \cdot w_{ij}$$

Where $IndA_j$ is the resulting activity of each node in the map due to indirect activation via Hebbian association, x_i is the pre-gaussian activity of unit i , in the other map and w_{ij} is the Hebbian weight between it and unit j in this map.

Equation 3: Gaussian direct or indirect activation of SOM units

$$y_i = e^{\frac{\|x - b_j\|^2}{2\sigma^2 n}}$$

Where y_i is the final activation of the i^{th} node in the map, β is the distance from node i to the winning unit (either direct or indirect), and n is the total number of nodes in the map. Note: units not within the neighborhood size are set to zero output activation, the neighborhood size and learning rate are logarithmically decreased.

Equation 4: Joint activation of SOM units

$$JoinT_i = (1 - \lambda)DirY_i + \lambda IndY_i$$

Where $JoinT_j$ is the final resulting activity of each node in the map due to the combination of direct and indirect activation, and λ is the activation mixture co-efficient (0.1).

Equation 5: SOM weight changes

$$\Delta w_{ij} = \alpha (v_i - w_{ij}) JoinT_i - \zeta (DirY_i - JoinT_i) (v_i - w_{ij})$$

Where w_{ij} is the weight between input j and unit i , α is the learning rate (0.1 – 0.0), and ζ is the inhibition rate (0.001 - 0.07).

Equation 6 Positive Hebbian learning (weight changes between maps)

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

$$\Delta w_{ij} = \alpha_{heb} DirA_i DirA_j$$

Where w_{ij} is the weight between node j and node i , α is the hebbian learning rate (0.01), $DirA_i$ is the initial direct activity of node i , in one map and $DirA_j$ is the initial direct activity of node j , in the other map.

The model is fairly robust to parameter changes, however we did 'tweak' the learning rates to obtain a best fit to the child data, BUT this same learning rate was then used across all experiments. Lower learning rates not only resulted in slower learning but consistently more errors, lowering the performance across all conditions in all experiments (with the exception of chance level performance), while retaining the same between condition differences. Similarly raising the learning rate not only speeded up learning but consistently raised performance across all conditions in all experiments, retaining the same between condition patterns (with the exception of perfect performance). Small SOMs lead to merging categories but other than that SOM size had no obvious effect.

References

- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(02), 105-124.
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences*, 11(7), 280-289.
- Bar, M. (2011). *Predictions in the brain: Using our past to generate a future*: Oxford University Press.
- Belpaeme, T., & Morse, A. (2012). Word and category learning in a continuous semantic domain: Comparing cross-situational and interactive learning. *Advances in Complex Systems*, 15(03n04).
- Burton, A. M., Young, A. W., Bruce, V., Johnston, R. A., & Ellis, A. W. (1991). Understanding covert recognition. *Cognition*, 39(2), 129-166.
- Chomsky, N. (1995). *The minimalist program* (Vol. 1765): Cambridge Univ Press.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303-306.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7(2-3), 195-225.
- Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological review*, 109(3), 545.
- Fodor, J. A. (1975). *The language and thought*: Harvard University Press.
- Fodor, J. A. (2010). *LOT 2: The language of thought revisited*: Oxford University Press.
- Frank, S. L. (2006). Strong systematicity in sentence processing by an Echo State Network *Artificial Neural Networks-ICANN 2006* (pp. 505-514): Springer.

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, 22(3-4), 455-479.

Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental psychology*, 40(4), 621.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning Words and Rules Abstract Knowledge of Word Order in Early Sentence Comprehension. *Psychological science*, 17(8), 684-691.

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, 17(01), 171-183.

Gonçalves, A., Abrantes, J., Saponaro, G., Jamone, L., & Bernardino, A. (2014). *Learning Intermediate Object Affordances: Towards the Development of a Tool Concept*. Paper presented at the IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob 2014), Genoa, Italy.

Hebb, D. O. (1952). *The organisation of behaviour: a neuropsychological theory*: Wiley.

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends in cognitive sciences*, 6(6), 242-247.

Hinton, G., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157.

Huttenlocher, P. R. (1994). Synaptogenesis in human cerebral cortex.

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148, 34.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, 88(5), 375.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, 317(5838), 631-631.

Morse, A. F., Belpaeme, T., Cangelosi, A., & Floccia, C. (2010). *Modeling U Shaped Performance Curves in Ongoing Development*. Paper presented at the Cognitive Science Conference.

Morse, A. F., DeGreeff, J., Belpaeme, T., & Cangelosi, A. (2010). Epigenetic Robotics Architecture (ERA). *IEEE Transactions on Autonomous Mental Development*, 2(4), 325-339.

Morse, A. F., Herrera, C., Clowes, R., Montebelli, A., & Ziemke, T. (2010). The Role of Robotic Modeling in Cognitive Science. *New Ideas In Psychology*.

Newell, A. (1990). *Unified theories of cognition*: Cambridge, MA: Harvard University Press.

Noe, A. (2004). *Action in perception*: the MIT Press.

O'regan, J. K. (2011). *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*: Oxford University Press.

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(05), 939-973.

Piaget, J., Cook, M., & Norton, W. (1952). *The origins of intelligence in children* (Vol. 8): International Universities Press New York.

Pinker, S. (1994). *The language instinct: The new science of language and mind* (Vol. 7529): Penguin UK.

Pulvermüller, F. (2002). *The neuroscience of language: on brain circuits of words and serial order*: Cambridge University Press.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576-582.

This is the author's accepted manuscript. The final published version of this work (the version of record) is published by Cognitive Science Society in Cognitive Science available at: 10.1111/cogs.12390. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

- Sisk, C. L., & Zehr, J. L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Frontiers in neuroendocrinology*, 26(3), 163-174.
- Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental science*, 14(1), 9-17.
- Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature neuroscience*, 6(3), 309-315.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381-382.
- Thelen, E., & Smith, L. B. (1998). Dynamic systems theories. *Handbook of child psychology*.
- Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in cognitive sciences*, 4(4), 156-163.
- Tong, M. H., Bickett, A. D., Christiansen, E. M., & Cottrell, G. W. (2007). Learning grammatical structure with echo state networks. *Neural Networks*, 20(3), 424-432.
- Twomey, K. E., Morse, A. F., Cangelosi, A., & Horst, J. S. (2014). Competition affects word learning in a developmental robotic system. In G. Westermann, P. Monaghan, T. K.E. & S. A (Eds.), *Proceedings of the 14th Neural Computation and Psychology Workshop (NCPW14)*. : World Scientific.
- Zukow-Goldring, P., & Arbib, M. A. (2007). Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing*, 70(13), 2181-2193.